

***ProtCV*, a tool for extracting, visualizing and validating protein clusters using mass spectra peak lists**

Stavroula Ventoura¹, Eugenia Giannopoulou², Elias Manolakos¹

¹*Department of Informatics and Telecommunications, University of Athens,* ²*University of Peloponese, Tripolis.*

In the post-genomic era there is a pressing need for algorithms and interactive tools which can provide effective analysis and visualization of the large-scale datasets resulting from high throughput proteomics experiments. In this project, we have designed and developed a novel interactive bioinformatics software tool, called ***Protein Clustering and Visualization (ProtCV)***, that can find applications in organizing, preprocessing and clustering sets of mass spectra peak-lists produced routinely in typical proteomics analysis studies. *ProtCV* can statistically preprocess and normalize peak-lists, create clusters of protein spots based on extracted peak-list patterns, visually summarize the clustering results using dendrograms, heat-maps, cluster sets and other advanced visualization techniques, quickly compare results obtained by different clustering algorithms, or by the same algorithm but with different user-defined parameters. Specifically *ProtCV* supports Hierarchical clustering (with different linkage methods and distance metrics) as well as k-means clustering (with different initialization methods etc). Formed protein clusters can be validated using the Silhouette method, Dunn's index, Davies-Bouldin index etc. in order to find the clustering that best captures the underlying structure of a peak lists dataset. *ProtCV* can be useful not only in assigning function to unidentified proteins but also in discovering protein networks and proteins that can jointly act as biomarkers in clinical proteomics applications.